SECTION 8

Uniform Laws of Large Numbers

For many estimation procedures, the first step in a proof of asymptotic normality is an argument to establish consistency. For estimators defined by some sort of maximization or minimization of a partial-sum process, consistency often follows by a simple continuity argument from an appropriate uniform law of large numbers. The maximal inequalities from Section 7 offer a painless means for establishing such uniformity results. This section will present both a uniform weak law of large numbers (convergence in probability) and a uniform strong law of large numbers (convergence almost surely).

The proof of the weak law will depend upon the following consequence of the first two lemmas from Section 3: *for every finite subset $\mathcal{F}$ of $\mathbb{R}^n$,*

$$\mathbb{P}_\sigma \max_{\mathcal{F}} |\boldsymbol{\sigma} \cdot \mathbf{f}| \le C \max_{\mathcal{F}} |\mathbf{f}|_2 \sqrt{2 + \log(\#\mathcal{F})}. \tag{8.1}$$

Here $\#\mathcal{F}$ denotes the number of vectors in $\mathcal{F}$, as usual, and C is a constant derived from the inequality between $\mathcal{L}^1$ and $\mathcal{L}^\Psi$ norms.

(8.2) THEOREM. *Let $f_1(\omega, t)$, $f_2(\omega, t)$, ... be independent processes with integrable envelopes $F_1(\omega)$, $F_2(\omega)$, ... If for each $\epsilon > 0$*

(i) there is a finite K such that

$$\frac{1}{n}\sum_{i\le n} \mathbb{P}F_i\{F_i > K\} < \epsilon \qquad \textit{for all } n,$$

(ii) $\log D_1(\epsilon|\mathbf{F}_n|, \mathcal{F}_{n\omega}) = o_p(n)$,

then

$$\frac{1}{n}\sup_t |S_n(\omega,t) - M_n(t)| \to 0 \qquad \textit{in probability.}$$

PROOF. Let us establish convergence in $\mathcal{L}^1$. Given $\epsilon > 0$, choose K as in assumption (i) and then define $f_i^*(\omega, t) = f_i(\omega, t)\{F_i(\omega) \le K\}$. The variables discarded by this truncation contribute less than 2ϵ:

$$\frac{1}{n}\mathbb{P}\sup_t\left|\sum_{i\le n}(f_i - f_i^*) - \mathbb{P}(f_i - f_i^*)\right| \le \frac{2}{n}\sum_{i\le n}\mathbb{P}F_i\{F_i > K\}.$$

For the remaining contributions from the $f_i^*(\omega, t)$ processes, invoke the symmetrization inequality from Theorem 2.2, with Φ equal to the identity function.

$$\frac{1}{n}\mathbb{P}\sup_t\left|\sum_{i\le n} f_i^* - \mathbb{P}f_i^*\right| \le \frac{2}{n}\mathbb{P}\mathbb{P}_\sigma \sup_{\mathcal{F}_{n\omega}}|\boldsymbol{\sigma}\cdot\mathbf{f}^*|.$$

Given ω, find a set $\mathcal{D}_{n\omega}$ of at most $M_n = D_1(\epsilon|\mathbf{F}_n|, \mathcal{F}_{n\omega})$ many points in $\mathcal{F}_{n\omega}$ that approximate each point of $\mathcal{F}_{n\omega}$ within an ℓ_1 distance of $\epsilon|\mathbf{F}_n|_1$. By assumption (ii), the random variables $\{\log M_n\}$ are of order $o_p(n)$. The expectation with respect to $\mathbb{P}_\sigma$ on the right-hand side of the last expression is less than

$$\frac{\epsilon}{n}|\mathbf{F}_n|_1 + \frac{1}{n}\mathbb{P}_\sigma \max_{\mathcal{D}_{n\omega}}|\boldsymbol{\sigma}\cdot\mathbf{f}^*|.$$

The first of these terms has a small expectation, because assumption (i) implies uniform boundedness of $\frac{1}{n}\mathbb{P}|\mathbf{F}_n|_1$. The second term is bounded by K. By virtue of inequality (8.1) it is also less than

$$\frac{C}{n}\max_{\mathcal{D}_{n\omega}}|\mathbf{f}^*|_2\sqrt{2 + 2\log M_n}.$$

The square root factor contributes at most $o_p(\sqrt{n})$ to this bound. The other factor is of order $O_p(\sqrt{n})$, because, for each point in $\mathcal{F}_{n\omega}$,

$$|\mathbf{f}^*|_2^2 = \sum_{i\le n} f_i^2\{F_i \le K\} \le K\sum_{i\le n} F_i.$$

A uniformly bounded sequence that converges in probability to zero also converges to zero in $\mathcal{L}^1$. □

When the processes $\{f_i(\omega, t)\}$ are identically distributed, the convergence in probability asserted by the theorem actually implies the stronger almost sure convergence, because the random variables

$$\frac{1}{n}\sup_t|S_n(\omega, t) - M_n(t)|$$

form a reversed submartingale. (Modulo measurability scruples, the argument for empirical processes given by Pollard (1984, page 22) carries over to the present context.) Without the assumption of identical distributions, we must strengthen the hypotheses of the theorem in order to deduce almost sure convergence. Manageability plus a second moment condition analogous to the requirement for the classical Kolmogorov strong law of large numbers will suffice. The stronger assumption about the packing numbers will not restrict our use of the resulting uniform strong law of large numbers for the applications in these notes; we will usually need manageability for other arguments leading to asymptotic normality.

(8.3) THEOREM. *Let $\{f_i(\omega, t) : t \in T\}$ be a sequence of independent processes that are manageable for their envelopes $\{F_i(\omega)\}$. If*

$$\sum_{i=1}^{\infty} \frac{\mathbb{P}F_i^2}{i^2} < \infty,$$

then

$$\frac{1}{n} \sup_t |S_n(\omega, t) - M_n(t)| \to 0 \qquad \textit{almost surely.}$$

PROOF. Define

$$\begin{aligned} f_i^*(\omega, t) &= f_i(\omega, t) - \mathbb{P}f_i(\cdot, t), \\ Z_{k,n}(\omega) &= \sup_t \left| \frac{f_k^*(\omega, t)}{k} + \cdots + \frac{f_n^*(\omega, t)}{n} \right| \qquad \text{for } k \le n, \\ B_k(\omega) &= \sup_{i,n \ge k} Z_{i,n}(\omega). \end{aligned}$$

By the triangle inequality

$$\begin{aligned} \sup_t \left| f_1^*(\omega, t) + \cdots + f_n^*(\omega, t) \right| &\le Z_{1,n}(\omega) + \cdots + Z_{n,n}(\omega) \\ &\le B_1(\omega) + \cdots + B_n(\omega). \end{aligned}$$

It therefore suffices to prove that $B_n \to 0$ almost surely.

From inequality (7.10) applied to the processes $f_i^*(\omega, t)/i$ instead of to $f_i(\omega, t)$, manageability implies existence of a constant C such that

$$\mathbb{P}Z_{k,n}^2 \le C \sum_{i=k}^{n} \frac{\mathbb{P}F_i^2}{i^2} \qquad \text{for } k \le n. \tag{8.4}$$

For fixed k, the random variables $Z_{k,n}$ for $n = k, k+1, \ldots$ form a submartingale. By Doob's (1953, page 317) inequality for nonnegative submartingales, for each m greater than k,

$$\mathbb{P} \max_{k \le n \le m} Z_{k,n}^2 \le 4\mathbb{P}Z_{k,m}^2.$$

Letting m tend to ∞, we deduce for each k that

$$\mathbb{P} \sup_{k \le n} Z_{k,n}^2 \le 4C \sum_{i=k}^{\infty} \frac{\mathbb{P}F_i^2}{i^2}.$$

The sum on the right-hand side converges to zero as $k \to \infty$. From the bound

$$B_k \le 2 \sup_{k \le n} Z_{k,n}$$

it follows that $\mathbb{P}B_k^2 \to 0$. Because $\{B_k\}$ is a decreasing sequence of random variables, it follows that $B_k \to 0$ almost surely, as required. □

REMARKS. Theorem 8.2 is based on Theorem 8.3 of Giné and Zinn (1984). They established both necessity and sufficiency for empirical processes with independent, identically distributed summands. The direct use of inequality (8.1) simplifies the argument, by avoiding an appeal to a chaining inequality. Except for the use of ℓ_2 packing numbers instead of ℓ_1 covering numbers, my proof is close to the proof of Theorem II.24 of Pollard (1984).

Theorem 8.3 is actually a special case of the Ito-Nisio Theorem (Jain and Marcus 1978, Section II.3). Zaman (1989) used a type 2 inequality analogous to (8.4) to reduce his proof of a uniform strong law of large numbers to the Ito-Nisio theorem. He imposed the same sort of moment condition as in Theorem 8.3.